DOCUMENT RESUME

ED 186 452 TM 800 139

AUTHOR Bayless, David L.: Nix, Charles W.
TITLE Competency Testing: Setting Educational Performance Standards for the Group.
PUB DATE 12 Jun 79
NOTE 41p.: Paper presented at the Annual Conference on Large-Scale Assessment (9th, Denver, CO, June 12, 1979).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Standards: Cutting Scores: Difficulty Level: Elementary Secondary Education: Geographic Regions: *Group Testing: *Individual Testing: Influences: *Minimum Competency Testing: Regional Programs: Sampling: State Programs: *Testing Problems: *Testing Programs: Test Items
IDENTIFIERS Texas

ABSTRACT

The merits and hazards of minimum competency testing for the individual student or for student groups are discussed. Types of groups which lend themselves to group application and some important factors in determining the parameters of a group are discussed. Ten critical issues related to minimum competency testing are identified: (1) scope of curriculum designated as minimum; (2) levels of expected proficiency; (3) determination of mastery; (4) performance criteria expressed as absolute standards or as achievement gains; (5) type of test; (6) test length and testing time; (7) costs of testing program and personnel time; (8) utilization of test results; (9) test security; and (10) protection of student rights. The relative advantages of individual versus group testing, and the option not to employ minimum competency testing at all are examined in terms of utility analysis and of the resultant gains and losses. The merits of criterion-referenced, objective-based, and norm-referenced testing are compared, especially the sampling problems involved. Two cases of application of minimum competency testing to groups are presented: one at the state level (Texas) and one for a multi-county region. (Texas' Region 12). (Author/GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED186452

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Competency Testing: Setting Educational Performance Standards for the Group

by

David L. Bayless
Senior Statistical Scientist
Research Triangle Institute

Charles W. Nix
Associate Commissioner for
Planning and Accreditation
Texas Education Agency

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

David Bayless

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

IM800 139

Paper presented at the
Ninth Annual Conference on Large-Scale Assessment
Denver, Colorado
June 12, 1979

ABSTRACT

In this paper the authors have attempted to throw some light upon the merits and hazards of implementing a minimum competency testing effort applied primarily to the individual student (MCT-I) or to groups of students (MCT-G). Types of groups which lend themselves to group-application, and some important factors in determining the parameters, of a group are discussed. Ten critical issues related to application of minimum competency testing to individuals or to groups are identified from the literature and from the authors' experience. Brief illustrations of the interactions of these issues in the decision-making process are provided. The relative advantages of MCT-I, MCT-G, and the option not to employ minimum competency testing at all are examined in terms of the utility analysis which must occur as a basis for deciding which of the options should be taken, and the trade-offs in terms of gains and losses for the educational agency are discussed. The relative merits of criterion-referenced, objective-based, and norm-referenced testing are discussed in terms of their use in applying minimum competency testing to individuals and to groups, especially in terms of some of the sampling problems involved. Finally, two cases of application of minimum competency testing to groups are presented: one at the state level and one for a multi-county region within a state.

TABLE OF CONTENTS

	Page
. Abstract	1
. Introduction	1
. Some Critical Issues	6
. Utility Analysis of the Minimum Competency Testing-Individual (MCT-I) and Minimum Competency Testing-Group (MCT-G) Options	14
. Two Cases of Group-Application of Minimum Competency Testing	25
. Summary	31
. References	34

COMPETENCY TESTING: SETTING EDUCATIONAL PERFORMANCE STANDARDS FOR THE GROUP

by

David L. Bayless, Research Triangle Institute
Charles W. Nix, Texas Education Agency

In its most generic sense, the term "minimum competency testing" refers to the process of defining a set of minimum competencies which a specified group of students — competency population[1/] — is expected to acquire through schooling, ascertaining whether or not such students have acquired the designated minimum competencies, and using the results of such assessment to increase the level of mastery of the minimum competencies by the competency population. Exactly what this means in operational terms has taken a variety of forms among the school districts, state departments of education, and other educational agencies which have moved into minimum competency testing. The basic unit of the competency population may be defined as the individual student, or it may be aggregations (groups) of students. In this paper these two distinct options or designs will be termed minimum competency testing applied to the individual (MCT-I) and minimum competency testing to groups (MCT-G).

Most of the theoretical development and application of minimum competency testing has focused upon the individual student as the competency population of interest, that is, designated minimum

---

[1/] The term competency population is used in this paper to mean a specified population of one or more students who are expected to master certain designated competencies.

competencies are expected to be acquired by each individual and a measurement system is established to ascertain whether or not each individual has mastered the designated minimum competencies. In many cases the student is required to attain a passing score on the test as a condition for grade promotion or graduation. Brickell (1978) states that "enthusiastic proponents believe they have found a way to force greater emphasis on the basic skills and to guarantee that the high school diploma will mean that certain skills and facts have been mastered by the (individual) graduate rather than that the graduate has 'put in his or her time'". In contrast to the focus upon application to individual students (MCT-I), there exist rich possibilities for the application of the minimum competency concept to groups of students (MCT-G), along with certain advantages which may be gained from MCT-G.

With the group-application, no requirement necessarily exists that an individual student demonstrate mastery of minimum competencies as a condition for begin promoted to the next grade or graduated from school; instead criteria are established as performance expectations for a group of students. In the group-application, emphasis is not placed upon the certification of competencies acquired by each individual; rather, the main emphasis is upon use of information about performance of groups to make changes (improvements) in the instructional delivery system which will increase the probability that larger proportions of individuals within the group(s) will master the minimum competencies as a result of increased effectiveness of instruction. This shift in focus from the individual to the group raises some

critical issues and places a different slant on some of the issues well-known in applying minimum competency testing to the individual. This paper will identify some of these issues, consider their relevance to individual-application and/or group-application, and illustrate some of the trade-offs or utilities (benefits and losses) which must be negotiated as policy makers set about deciding which type of application, MCT-I or MCT-G, best fits the requirements of their particular situation.

An essential element of a minimum competency testing program that uses the group as the competency population is the definition of the group. Theoretically, a group is defined to be two or more individuals with some characteristic(s) in common who belong to the competency population that is being instructed and assessed; it is assumed that a statement has been developed that unequivocally defines which individuals belong to the competency population, such as all of the eleventh graders attending public schools in a given state, or all pupils enrolled in a given elementary school. Table 1 contains some practical illustrations of types of groups that may be considered in the MCT-G design.

Table 1. Illustrations of Groups

Group Name	Description of the Group
1. Campus (school)	All members of the competency population enrolled in a campus or school building.
2. School district	All members of the competency population in a given school district.

Table 1 (continued)

3. Region	All members of the competency population attending schools that are located in a specified geographical area defined by a collection of counties or school districts within a state.
4. State	All students attending the public schools in a state.
5. Demographic characteristics	All members of the competency population having a common demographic characteristic, such as sex, ethnicity, family income, parents' education, etc. The members of such a demographic group could be defined to be within a school, school district, region, or state.

The classroom as a group is not included as an illustration in this paper since the interest here is upon large-scale policy concerns and not upon the performance of individual teachers. Table 1 illustrates a range of groups that may be designated as competency populations from a campus to a whole "state", as well as other configurations such as members of an ethnic group of a school or state.

Which group is of greatest utility for a given application of minimum competency testing depends upon several practical factors that characterize the competency testing program being planned. Such factors include:

1. Size of the group. For some MCT-G applications, consideration should be given to the number of individuals in the group. If the group is too small it may be difficult to set reasonable standards for the group to achieve; too large a group

(such as large state) may be relatively useless as a competency population in that the policy makers may not have sufficient knowledge and control of the instructional practices that must be changed, i.e., repaired or replaced, to improve student performance.

2. Current level of student performance. This factor should be considered in establishing a group as the competency population. Known differences in performance levels of sub-populations with different demographic characteristics may be one of the most cogent bases for the designation of various competency populations.

3. Accessibility to policy information and direction. Different natural groups vary in the degree to which program conditions can be known to policy makers and can effectively be changed by the policy makers who control them. In some educational systems, the school district might better fit the above conditions, such as a small school district where the superintendent knows in detail what is going on in the schools of the district. In other educational systems, such as a large urban school district, the campus or school building might better serve as the competency population.

The purpose of this paper is to identify and consider several issues encountered in the application of MCT-G as contrasted with MCT-I and to analyze the utilities or trade-offs between setting

performance standards for the group or for the individual as the competency population for a minimum competency testing program. The next section outlines some important issues that should be considered in deciding whether the group or the individual should be established as the competency population. The following section discusses the trade-offs which must be made in deciding between the two options by describing comparative advantages and disadvantages among MCT-I, MCT-G, and no minimum competency standard at all.

The analysis is made of the MCT-I and MCT-G options when a norm-referenced, criterion-referenced, or objective based measurement methodology is used in assessment. Some real-life situations are described in which minimum competency testing has been applied to groups. The final section summarizes major points addressed in this paper.

Some Critical Issues

There are several important dimensions which should be considered in reaching a decision on whether individual-application (MCT-I) or group-application (MCT-G) of minimum competency testing is more appropriate for a given organizational entity. These issues include:

1. Scope and content of the curriculum designated as basic, essential, or minimum for students to master by the time they complete school. How much of the total curriculum should be covered in this designation? Should it be restricted to rudimentary elements of language and numbers? Or should it include values, attitudes, and higher order thinking processes as well? Should the minimum competencies

be limited to recall-type content? Are there content areas that are expected to be mastered by only some portion of the total student population? Madaus and Airasian (1977) suggest that some of the higher level skills of analysis, synthesis, and evaluation may not be teachable to all students, and that the home and other institutions have a major role along with the school in teaching values and attitudes.

2. Minimum levels of proficiency (mastery) expected to be attained by students in the designated basic or essential content areas. What proportion of the student population can the educational agency afford not to meet the established proficiency levels? Will the cut-offs be dichotomous (pass/fail)? Or will there be one or more intermediate classifications between pass and fail (mastery/near mastery/non-mastery)?

3. Determination of satisfactory performance on the tests. Will a specified percentage of items answered correctly constitute a "passing" score on the tests? Or will a passing score be defined in terms of demonstrated mastery of a specified number of the competencies measured? Or, since the curriculum material upon which the tests is based are termed "minimum" competencies, should the criterion for satisfactory performance be defined as demonstrated mastery of all the minimum competencies by all the students tested? How will complexity of curriculum

content and difficulty of test items be taken into account in deciding upon criteria for satisfactory performance that are at once fair, respectable, and useful?

4. Performance criteria expressed in absolute standards or relative gains. Shall the competency population be expected to attain a uniform performance level which is standard for everyone? Or shall performance expectations be based upon entry-level performance of the competency population? Will there be a single standard of mastery established for all students in the population, regardless of their educational backgrounds or other related factors? Will there be different cut-off scores for different campuses in a district?

5. Psychometric methods used to ascertain whether members of the population have mastered minimum competencies. Will conventional methods of multiple-choice or other "responding-mode" measures be relied upon to test mastery? Or will other more direct performance-based measurement modes be used, especially those which call for "initiating-mode" responses on the parts of students? What about the relative costs of these two measurement approaches, in terms of both money and student time?

6. Parsimony in balancing amount of testing time for students with amount of student performance data necessary to make sound decisions. How can student instructional time be protected and student testing time be minimized, while at the same time measuring mastery of a significant segment of the curriculum in reliable and valid ways? Must every student be tested on every competency? Or, will representative samples suffice? Will matrix sampling provide the information required? Is it feasible to administer a short screening test to all students to identify "moderate-to-high" masterers, followed by more detailed testing of low performers?

7. Schedule and personnel to administer the tests. Who shall administer the tests to the student population? Can the educational agency afford the cost of external administrators if the population to be tested is large? Do teachers tend to place more credence in the results and to use the information to evaluate and improve instructional practice if they administer the tests to their students? Should the designated minimum competencies be measured at a single sitting for a given competency population? Or can teachers administer items from the tests in synchronization with individually-paced instruction, keeping individual student performance records?

8. Maximizing the use of test results in deciding how best to re-cycle instruction to increase the probability of student mastery of minimum competencies. Is primary emphasis upon ascertaining whether or not students have met performance standards? Or is primary emphasis upon gaining information that will enable practitioners to change the probabilities for producing mastery? Should non-masters simply be re-cycled through the same instruction repeatedly until they master the minimum competencies? Or should efforts be made to vary instructional procedures to improve results? How can the test results obtained for high school students be used to improve instruction for students in the early grades? Is the individual student primarily accountable for producing mastery? Or is the instructional delivery system primarily accountable for finding instructional patterns that work for different students?

9. Security of material contained in tests. Can the same tests be used for different student populations on subsequent occasions? In developing new (equivalent) forms of the same test, can some items from earlier forms be brought forward? Is there available sufficient test-making expertise to develop a broad base of test items? How important is it to equate the level of difficulty at which cut-off scores are established for equivalent forms of a test covering

the same minimum competencies and administered to successive groups of students? What is the cost advantage of not being required to develop an entirely new form for each successive administration of a test? How can teachers and students be informed of the minimum competencies to be measured without giving them the opportunity to practice the items on the actual test?

10. Protection of the rights of students. How can the educational agency ensure that an individual student's scores are not released to unauthorized sources? How will educational agencies handle legal charges against decisions to withhold benefits from a student (promotion, graduation) based on test results? How will claims be answered that cultural bias in tests produces false negatives? Will these pressures lead to by-passing of standards by lowering cut-off scores, permitting promotion or graduation with parental waiver, or other circumventions?

The breadth and complexity of the issues outlined above preclude any attempt in this paper to cover comprehensively the many ways in which the issues, and the trade-offs involved, are interwoven into the process of deciding whether to apply minimum competency testing to individuals or to groups. A few brief observations will illustrate the ways in which these issues underlie the trade-offs involved, as well as some of the ways the issues themselves interface in the decision-making process:

- If every student is to be tested on every competency, it appears that the scope of competencies designated as minimum will have to be more limited. If different representative samples of students are to be tested on different competencies, then a broader scope of competencies can be covered without exposing students to inordinate amounts of testing or attempting to reach conclusions about student mastery on the basis of inadequate numbers of test items taken.
- If decisions consequential to the welfare of the individual student (promotion, graduation) are to be based on measurement results, then it appears prudent to limit minimum competencies to the rudimentary areas of language and mathematics performance, where the state of the art of measurement is better developed and "safer". If decisions are to be made only with respect to groups, then some of the hazards of measuring competencies in "softer" areas of the curriculum are greatly reduced or eliminated.
- Trichotomous criteria for adequacy of mastery of minimum competencies appear to be infeasible for individual applications where promotion and graduation decisions are to be made on the basis of test results. With applications to groups, however, a "near-mastery" category may be useful in judging the criticality of the measured discrepancy between actual and expected performance.

- Conventional multiple-choice test items are the most readily available and least expensive to use in testing every student in a large competency population. If small representative samples of students are tested, and the results generalized to the larger competency population, it is more feasible to use some of the more costly and time-consuming (yet more useful) testing methods which directly measure complex student behaviors such as written composition.
- In many cases it will be judged necessary to have teachers administer the competency tests to their own students. To the extent that teachers' perception of threat to the welfare of their students or to themselves can be reduced, policy makers will be able to expect more objective and uniform administration of the tests. In the group-application, participation of teachers in administering the tests can be brought into the service of the same teachers' participation in using test results in making decisions to improve instructional practice. Some of the strict measures to maintain test security may also be reduced.
- Building alternative forms of a test is expensive. It is also hazardous with respect to the assumed equivalence of the content and difficulty levels of alternative forms. This hazard is great enough where alternative forms are assumed to be equivalent only in terms of a global score for the entire test; where equivalency is desired for the

array of competencies which constitute the test, the problem is greatly magnified. Group-application offers greater possibilities for combining common items and new items on equivalent forms of a test and thus providing some degree of relief on this issue.

- Protection of the rights of students can be better safeguarded if there exist no records of test scores for individual students. Individual-application of minimum competency testing requires that records be maintained and used for individual students. Group-application affords an opportunity to deal with this problem at its source, that is, by not generating information about individual student performance.

Utility Analysis of the MCT-I and MCT-G Options

The issues dealt with in the foregoing section weave in and out of any analysis of the relative utility and merits of individual-application vs. group-application of minimum competency testing to any real-life situation. It should be pointed out that no claim is made that policy makers either adopt MCT-I or MCT-G for their situation. These two positions are viewed as "pure" poles for purposes of comparison and discussion. Any real decision to adopt minimum competency testing will very likely combine desirable elements of both positions to fit situational values and requirements. What remains, however, is the likelihood that any real application of minimum competency testing will be preponderantly in the direction of

individual-application or group-application. Additionally, no claim is made that MCT-I or MCT-G is inherently superior in the abstract; in almost all real cases, policy makers cannot have all the benefits (or avoid all the difficulties) of both individual-application and group-application, and certain trade-off decisions must be taken with respect to the utility of each and the prevailing values and purposes of the educational agency. Therefore, this section is not intended as an argument in favor of the application of minimum competency testing to groups. It is intended to display some of the utility analysis and trade-off decisions which will occur in selecting the mode of implementing minimum competency testing which provides the greatest gains and the fewest losses.

The utilities of competency testing when educational performance standards are set for either the individual or for groups are provided in Table 2. For completeness, consideration is also given to the advantages and disadvantages of setting no performance standard at all, that is, when competency testing is not implemented. The statements in Table 2 are from a variety of sources and references that have been published in the literature along with the authors' observations from practical experience in the planning, design, and implementation of national, state, and local assessments, research studies, and program evaluations. The statements are considered to be representative and reasonably comprehensive for the purpose of this utility analysis, but do not purport to be exhaustive of all relevant factors that could be considered in analyzing the MCT-G and MCT-I options.[2/]

[2/] See the following references: AASA Critical Issues Report (1978); Brickell (1978); Conaway (1977), Gallagher (1979); Madaus and Airasian (1977); and Meskauskas (1976).

Table 2

Advantages and disadvantages of the MCT-I and MCT-G options and the option of not setting standards

	Advantages	Disadvantages
MCT-G: Minimum Competency Testing Applied to the Group	. Permits inclusion of broader curriculum content in designated minimum competencies. . Identifies groups with low basic skills and facilitates remediation. . Results can be used to repair or replace program weaknesses. . Facilitates setting of group standards in terms of expected gain scores over time. . Permits the use of sampling which can be cheaper and provide quality data. . Allows for a wide range of item difficulty in the measurement design. . Individual results are anonymous.	. Weaknesses of individuals are not identified and hence not remediated. . Permits more non-masterers to be certified as false positives. . Requires larger-scale remedial actions, more difficult to initiate.
MCT-I: Minimum Competency Testing Applied to the Individual	. Identifies individuals with low basic skills and facilitates remediation. . Increases certitude that individuals meet standards prior to certification. . Decreases social promotion, if standard is high enough.	. Requires long tests with many items to ensure fair coverage of competencies. . Testing time wastes instruc-time of bright students. . Individual results are not anonymous. . May contribute to an individual's dropping out of school. . Encourages teachers to focus on the competencies of the test, narrowing the learning opportunities offerred. . Risks misclassification of certain individuals, consequently being unfair and inconsistent. . Scope of school outcomes that can be measured is narrow. . Requires more effort to keep tests secure.

-- continued --

Table 2 (Continued)

	Advantages	Disadvantages
No Minimum Competency Testing	. Avoids controversy. . No individuals are misclassified. . Saves resources that can be allocated to instruction.	. Permits social promotion. . Some individuals will not receive effective instruction in basic skills. . Limits information about individual or group achievement as basis of improving instruction.

Some critics of competency testing for the individual, Glass (1978 a and b), maintain that MCT-I should not be done at all, because it is unfair to the student in that the performance standards are decided upon in an "arbitrary and capricious way". The authors agree with Glass that the progress over time performance standard he suggests has considerable merit for MCT-G design, but cannot agree with the position that the MCT-I design should not be used at all.

Table 2 is summarized by classifying advantages of the MCT-G design and disadvantages of the MCT-I design into two major categories: methodological concerns and policy implications for results. The methodological advantages of the MCT-G design are:

- Permits the use of sampling which can provide quality data at a lesser cost.
- Allows for a wide range of item difficulty in the measurement design.
- Permits inclusion of broader curriculum content in designated minimum competencies.

The advantages of MCT-G design from a policy implication viewpoint are:

- Facilitates setting of group standards in terms of expected or targeted gain scores over time.
- Results can be used to repair and replace ineffective instruction.
- Identifies groups with "low" basic skills and facilitates remediation.

The methodological disadvantages of the MCT-I design are:

- Scope of school outcomes that can be measured is relatively narrow.
- Requires more effort to keep tests secure.
- Individual results are not anonymous.

The disadvantages of the MCT-I design from a policy implication viewpoint are:

- Encourages teachers to focus inordinately on the competencies measured, narrowing the learning opportunities offered.
- May contribute to an individual's dropping out of school.
- Testing time waste instruction time of able students.
- Risks misclassification of certain individuals, consequently being unfair and inconsistent.
- Can be unfair to individuals whose performance is below the standard because their schooling did not cover the competencies measured on the test or they were "poorly instructed" on competencies covered by the test.

It is not the purpose of this paper to study in detail the differences among the criterion-referenced testing (CRT), objective based testing (OBT), and norm-referenced testing (NRT). References exist that delineate the advantages and disadvantages of using such methods in implementing MCT.[3/] The purpose for this paper is to comment on the relevance of these measurement methods to the MCT-I and MCT-G designs.

3/ See the references of Hambleton, et al (1978), and Mehrens and Ebel (1979).

Table 3 provides a description of the types of sampling that are applicable for the CRT, OBT, and NRT measurement methods for both the sampling of persons and items. Probability and nonprobability sampling methods are classified. By probability sampling it is meant that each individual in the competency population has a positive chance of being included into the sample. By nonprobability sampling, it is meant that each individual is selected into the sample using a purposeful or nonrandom method of selection.

Probability sampling methods are not limited strictly to simple random sampling, but also include the more "complex" sampling designs such as stratified random sampling, stratified multi-stage or cluster sampling, etc. Probability sampling can be less costly in that it requires less testing time, smaller number of students to be tested and, consequently, a smaller number of schools that must be involved in the administration of the tests. Sampling permits better control of both sampling and nonsampling errors of the results of the MCT-G design, e.g., sampling errors through the sample design and nonsampling errors through the survey or data collection design. Nonsampling errors are controlled by the development and implementation of conditions in the measurement aspects of the testing, e.g., use of trained and qualified test administrators to collect the data from the students tested. The implementation of methods to control nonsampling errors avoid various threats of bias to the measurement aspects of an MCT in that the data collected from each individual tested can be of a higher quality and that teachers and other school personnel are precluded from teaching the test to the student before and/or during the administration of the test.

Table 3

Types of person and item sampling (probability versus nonprobability) required for Criterion-Referenced Testing (CRT), Objective Based Testing (OBT), and Norm Referenced Testing (NRT)

	Criterion-Referenced Testing	Objective Based Testing	Norm-Referenced Testing
Person	Probability	Probability	Probability
Item	Probability	Nonprobability	Nonprobability

It is essential with all probability sampling methods that the individual results be weighted in accordance with the reciprocal of the overall probability of the individuals being selected in the assessment sample. This aspect should be considered in the analysis of the data by those that use latent trait models (Rasch model) to construct minimal competency tests. Examinee weights are required for most applications of the latent trait model since it is not valid to assume that the sample of individuals is simple random or that the responses between individuals are independent of each other. Several empirical studies demonstrate that the individual responses across individuals are not independent for sample survey design commonly used in published survey research.[4/] Educational researchers using the latent trait model would be wise to study this literature for its potential application in the analysis of data when a latent trait model is assumed.

The nonprobability sampling methods used for the selection of items for the OBT and NRT measurement methods are equivalent to what social researchers call quota sampling. The use of such a methodology for item sampling in OBT and NRT applications could be of value to the quality of the norm-referenced or objective based tests in that quotas can be established relative to the difficulty of the items of the tests, e.g., easy, medium, and difficult. In addition, quotas can be set to cover the various content sub-populations of a given objective or competency that is being measured. For the NRT measurement method, different quotas are set for the item

---

[4/] See the monograph by Frankel (1971) for a reference.

sampling procedures. Since the major use of NRT results is to reference NRT scores to a person population so that the individuals can be rank-ordered or normed on a given test scale, the special quotas are set for items that are around fifty percent item difficulty in order to provide a wide range of test score variability.

Table 4 classifies the range of item difficulty and scope of item content into two categories, narrow and wide. Table 4 also classifies the generalizability of the results into two categories, poor and excellent, for the MCT-I and MCT-G designs and three measurement methods.

The major conclusion to be drawn from Table 4 is that for most applications, the MCT-G design makes it possible to include a wide scope of item content with a wide range of item difficulty in the competency test. The CRT measurement method used in connection with the MCT-G design has the additional advantages:

- It permits sampling of items from a large number of domains
- It promotes balance among testing time, number of objectives measured, and instructional time consumed.
- Samples of items included in the competency tests are representative of the item population defined by domain specification.

A disadvantage of criterion-referenced testing is that the domain specifications are not clear to item writers and potential users of test results. Thus, the validity of the generalizability of the test results to the content population may be questioned. A major weakness of the NRT measurement method for the MCT-G design is that

Table 4

Range of item difficulty, scope of item content, and generalizability of results by measurement methods (CRT, OBT, and NRT) and by MCT-G and MCT-I designs

	Measurement method	Range of item difficulty	Scope of item content	Generalizability of individual/group results to item population defined by domain specifications
MCT-G: Minimum Competency Testing Applied to the Group	CRT	Wide	Wide	Excellent
	OBT	Wide	Wide	Excellent
	NRT	Wide	Narrow	Poor
MCT-I: Minimum Competency Testing Applied to the Individual	CRT	Narrow	Narrow	Excellent
	OBT	Narrow	Narrow	Excellent
	NRT	Wide	Narrow	Poor

the test results often do not provide information on what students can and cannot do, with respect to the designated minimum competencies. However, NRT data can be useful in an MCT-G design in that NRT results can be of supplemental use in setting reasonable standards of educational performance for the group in conjunction with other data.[5/]

Two Cases of Group-Application of Minimum Competency Testing

There are a number of actual examples of the application of the minimum competency testing concept to groups of students. One such example exists at the state level in Texas. Several years ago the Texas State Board of Education established as one of its statewide priorities the improvement of achievement in reading and mathematics among students in the public schools of the State. For this priority, the State Board set as an objective that

> Consistent with the Goals for Public School Education in Texas, by 1980, students completing high school should demonstrate the attainment of the essential knowledge, skills, and competencies in reading and mathematics.

This objective did not specify that <u>every</u> student should demonstrate attainment of these "essential" (synonyous with "minimum") competencies, but that <u>all</u> students as a statewide group should do so. As the State Board discussed the meaning of this objective, it became increasingly clear that relying upon a standardized norm-referenced test to define

---

[5/] The paper by Conaway (1976) is an excellent reference on this point.

these essential competencies, and using any one of the composite scores provided by such norm-referenced tests, would not serve as an adequate definition or measure of the basic competencies the Board had in mind. The staff of the State Department of Education was directed to develop a definition of these essential reading and mathematics competencies and to devise a reliable and accurate method of measuring and reporting their attainment by students in the State.

Through regional meetings involving educators, lay citizens, and students, lists of essential competencies in reading and mathematics were generated for two levels of schooling;

- exit-level student outcomes were designed to define those competencies in reading and mathematics expected to be acquired by all students by the time they have completed school, and were intended to reflect those minimum or essential competencies necessary to function in adult life.
- mid-point level outcomes were designed to indicate the level of mastery of these essential competencies expected to be attained by students half-way through the schooling process (approximately grade 6), as check-points on the way to acquisition of the exit-level competencies.

Since these mid-point and exit-level student learning outcomes were labeled as "essential" outcomes, it was the expectation that all students would master them at an appropriate time in their schooling. The concept "all", however, was not viewed as absolute, since

from a practical standpoint there will always be a small percentage of students in any competency population who, for various reasons, will not master the intended competencies.

After examining a wide range of currently available norm-referenced and criterion-referenced tests, the staff determined that no existing tests could adequately measure student attainment of these essential reading and mathematics competencies in the way the State Board intended. Development of a set of objective-based tests was undertaken by the State Department of Education, with the assistance of measurement personnel from a number of school districts in the State. The Educational Testing Service assisted in generating and field-testing items for some of the tests. Personnel from the National Assessment of Educational Progress provided technical advice at various points during the development of the tests. The result was a series of four tests based upon, and designed to measure achievement of, selected essential reading and mathematics competencies among 6th and 11th grade students.

In the spring of 1978 these tests were administered to representative samples of 6th and 11th grade students in Texas schools. No "pass/fail" scores were established as a means of interpreting the results, since it was desired that all students demonstrate mastery of each of the essential competencies measured. Results were analyzed, and a p-value or percent mastery of 90 or greater for an objective was established, for practical purposes, as an indication that "all" students in the competency population had mastered the competency. A p-value of less than 75 for an objective was taken as an indication

that a large gap existed between expected and actual performance of the competency population, and these gaps were reported as high priority student learning needs.

Since no effort was made to assess any individual student's performance or to prescribe specific remediation for him or her through this assessment, the emphasis in follow-up actions fell upon what could be done to improve instructional practices and, consequently, percentages of students mastering those essential outcomes for which group performance had been poor. Several such actions were taken:

- The State Board of Education requested that the State legislature appropriate funds to be used in developing and implementing instructional practices and inservice training of teachers that would produce better student mastery of the essential outcomes where priority needs were found.
- The statewide priorities for awarding of developmental grants under ESEA, Title IV-C were modified to focus upon the priority student learning needs identified.
- The seven largest urban school districts identified the achievement discrepancies found among urban students in the competency population (thus defining an urban sub-population as their target group), developed strategies to bring about the necessary improvement of instruction in urban schools, and requested assistance from the legislature.

- Since the results were analyzed and reported for each of the 20 regions of the State as well as for the State as a whole, each regional education service center was encouraged to gear its services to its constituent school districts in ways that would strengthen local school instructional programs that impact the identified discrepancies in student mastery of essential competencies.

None of the analysis or reporting of test results, nor any of the subsequent follow-up activities, revolved around the individual student. All were centered upon the performance of groups of students: statewide grade-level populations and demographic sub-populations, as well as geographically defined regional sub-populations. In no case was the score of any individual student interpreted or reported. Identification of individuals was destroyed after administration of the tests. No score was designated as passing for any of the tests, and no pass/fail judgments were made about the performance of groups of students, except for the expressed intention to increase over time the number of students mastering those essential outcomes for which p-values had been less than 90.

A second example is found in a developmental project which has been conducted over the past three years by the Educational Service Center in Region 12, located in Waco, Texas. The Education Service Center undertook to develop a system to measure and improve student achievement along the same essential reading and mathematics outcomes addressed in the statewide effort described above. Region 12 contracted

with the ORBIT System of the California Test Bureau to develop objective-based tests of these essential competencies for students in grades 6 and 12. As in the statewide effort, it was taken as an expectation that, since these learning outcomes in reading and mathematics were styled as "essential" competencies, they should be mastered by all students. The tests were administered to students in approximately 50 school districts in Region 12. The education service center interpreted a p-value of less than 100 for any competency to be a discrepancy, and selected highest priority needs on the basis of the magnitude of the difference between 100 and the actual p-values obtained for the various competencies measured.

Several actions were taken to contribute toward strengthening those instructional practices in schools that bear upon the identified student achievement discrepancies:

- Curriculum consultants on the staff of the Region 12 Education Service Center used the identified student learning needs as the focus for inservice activities planned for school districts for the coming year. This resulted in a number of substantive changes in emphasis and operational strategies for inservice.
- The education service center's instructional media available for loan to school districts and the nature of technical assistance in curriculum planning provided to school districts are being reviewed and modified to give increased attention to those essential competencies where priority student needs were found.

- Proven instructional practices which show promise for improving student performance on the essential competencies with low p-values are being sought for dissemination to local school curriculum planners.

As in the statewide application, the Region 12 Education Service Center set no pass/fail scores for the tests, and no individual student scores are being reported or used for follow-up action in this project. Plans for the next phase of the project call for emphasizing the use of school district aggregated results, and campus aggregated results, in bringing more specific information to bear upon planning for instructional improvements in the institutions that provide instruction directly to students. Emphasis will continue to be upon aggregation of test results for competency populations defined in terms of region, school district, and school campus. Any application of results to individual students is incidental and remains solely at the discretion of the participating school districts.

Summary

In this paper the authors have attempted to throw some light upon the merits and hazards of implementing a minimum competency testing effort applied primarily to the individual student (MCT-I) or to groups of students (MCT-G). Types of groups which lend themselves to group-application, and some important factors in determining the parameters of a group, were discussed. Ten critical issues related to application of minimum competency testing to individuals or to groups were identified from the literature and from the authors' experience.

Brief illustrations of the interactions of these issues in the decision-making process were provided. The relative advantages of MCT-I, MCT-G, and the option not to employ minimum competency testing at all were examined in terms of the utility analysis which must occur as a basis for deciding which of the options should be taken, and the trade-offs in terms of gains and losses for the educational agency were discussed. The relative merits of criterion-referenced, objective-based, and norm-referenced testing were discussed in terms of their use in applying minimum competency testing to individuals and to groups, especially in terms of some of the sampling problems involved. Finally, two cases of application of minimum competency testing to groups were presented: one at the state level and one for a multi-county region within a state.

The authors' conclude that (1) minimum competency testing is probably never implemented purely in an individual-application or a group-application in the real world, but more commonly as a combination of elements of the two options, usually with a preponderance of one or the other, and (2) there is no valid argument that MCT-I is inherently better than MCT-G, or vice versa, but each has mutually exclusive advantages and disadvantages which policy makers must weight in relation to the values and purposes of their particular situation.

In the final analysis, one large, overriding trade-off must be made. In selecting the individual-application of minimum competency testing (MCT-I) the educational agency gains certitude in ascertaining that every individual student has met established standards before certification in the form of grade promotion or graduation is awarded, at the expense of diminished emphasis upon obtaining product

information that will permit effective identification of weaknesses in the instructional process. In opting for the group-application (MCT-G), the educational agency gains in focusing its attention upon information useful in deciding where the instructional process is weak and how it can be strengthened, at the expense of certitude in ascertaining whether individual students have met standards. The educational agency cannot have all these advantages at the same time, and a rational choice must be made within the context of the institution's values. It is the conviction of the authors, however, that in many cases there is more to be gained, and fewer important losses, in choosing the group-application approach when minimum competency testing is to be implemented.

REFERENCES

American Association of School Adminstrators Critical Issues Report (1978). The Competency Movement, Education News Service, Sacramento, California.

Brickell, Henry M. (1978). Minimum Competency Testing, Education Commission of the States, Report No. 110.

Conaway, L. E. (1977). Setting standards in competency-based education: some current practices and concerns. Paper presented at the annual meeting of NCME, New York.

Frankel, Martin R. (1971). Inference from Survey Samples: An Empirical Investigation, University of Michigan, Ann Arbor, Michigan.

Gallagher, James. J. (1979). Minimum Competency: The Setting of Educational Standards, Educational Evaluation and Policy Analysis, Vol. 1, No. 1, 62-67.

Glass, G. V. (1978). Standards and criteria. Journal of Educational Measurement, 15, 237-261. (a).

Glass, G. V. (1978). Minimum competence and incompetence in Florida. Phi Delta Kappa, 59, No. 9, 602-605. (b).

Hambleton, R. K., Swaminathan, H., Algina, J., and Coulson, D. B. (1978). Criterion-referenced testing and measurement: a review of technical issues and developments. Review of Educational Research, 48, 1-47.

Hambleton, Ronald K., Swaminathan, Hariharan, et al., (1978), Developments in latent trait theory: models, technical issues, and applications, Review of Educational Research, Vol. 48, No. 4., 467-510.

Madaus, George F., and Airasian, Peter W., (1977). Issues in evaluating student outcomes in competency-based graduation programs, Journal of Research and Development in Education, Vol. 10, No. 3, 80-91.

Mehrens, William A., and Ebel, Robert L. (1979). Some comments on criterion-referenced and norm-referenced achievements tests, NCME Measurement in Education, Vol. 10, No. 1.

Meskauskas, J. A. (1976). Evaluation models for criterion-referenced testing: views regarding mastery and standard-testing. Review of Educational Research, Vol. 46, 133-158.